The bromodomain: a conserved sequence found in human, *Drosophila* and yeast proteins

Susan R.Haynes, Catherine Dollard[1], Fred Winston[1], Stephan Beck[2], John Trowsdale[2] and Igor B.Dawid

Laboratory of Molecular Genetics, NICHD, NIH, Bethesda, MD 20892, [1]Department of Genetics, Harvard Medical School, Boston, MA 02115, USA and [2]Imperial Cancer Research Fund, London WC2A 3PX, UK

Submitted April 2, 1992

GenBank accession nos*

Identification of conserved domains or motifs in proteins may aid in the localization and analysis of important structural and functional regions. We report here a protein sequence motif, called the bromodomain (1), that has been found in six genes from humans (CCG1 and RING3), *Drosophila* (*fsh* and *brm*), and yeast (SPT7 and SNF2). The *fsh* and *brm* genes are required maternally for proper expression of certain homeotic genes (1,2). The *SPT7* and *SNF2* genes of *Saccharomyces cerevisiae* encode transcriptional activators (3,4). The *SNF2* and *brm* proteins have extensive sequence homology (1). It is not clear whether the two human genes are involved in processes. CCG1 is a DNA-binding protein that complements temperature sensitive mutations that cause cell cycle arrest in G1 (5). The RING3 gene is a newly identified human gene of unknown function, mapping to the class II region of the human major histocompatibility locus, that has substantial homology to the *fsh* gene (6).

Each of these proteins has one or two copies of the 61–63 amino acid bromodomain (Figure 1). The sequence identity is highest (~80%) between the corresponding *fsh* and RING3 repeats. More typically, any two motifs show 25–40% identity, with 50–60% identity between two repeats within the same protein. There are seven invariant residues, four of which are aromatic amino acids, and numerous conservative substitutions. The location of the motif(s) within the individual proteins is variable, and two motifs may be present in tandem or separated by sequences unrelated to the motif.

Secondary structure prediction methods were applied to the bromodomain and revealed two strongly predicted amphipathic α helices followed by reverse turns. Four of the invariant residues are located within the turns, and the other three are located among four highly conserved proline residues at the N terminus.

The functional significance of the bromodomain is unknown, and experiments with the two yeast proteins indicate that it may be dispensable in some cases (6; P.Tan and F.W., unpublished data). However, the widespread conservation of the bromodomain suggests that it is important for some aspects of protein function. We speculate that the hydrophobic surfaces of the helices and the invariant hydrophobic residues could serve as sites of intramolecular or intermolecular protein–protein interaction. Such interactions could influence the assembly or activity of multicomponent complexes involved in transcription activation or other cellular processes.

REFERENCES

1. Tamkun,J.W., Deuring,R., Scott,M.P., Kissinger,M., Pattatucci,A.M., Kaufman,T.C. and Kennison,J.A. (1992) *Cell* **68**, 561–572.
2. Huang,D.-H. and Dawid,I.B. (1990) *New Biologist* **2**, 163–170.
3. Winston,F., Dollard,C., Malone,E.A., Clare,J., Kapakos,J.G., Farabough,P. and Minehart,P.L. (1987) *Genetics* **115**, 649–656.
4. Laurent,B.C., Treitel,M.A. and Carlson,M. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 2687–2691.
5. Sekiguchi,T., Nohiro,Y., Nakamura,Y., Hisamoto,N. and Nishimoto,T. (1991) *Mol. Cell. Biol.* **11**, 3317–3325.
6. Beck,S., Hanson,I., Kelly,A., Pappin,D.J. and Trowsdale,J. (1992) *DNA Sequence*, **2**, 203–210.
7. Nishikawa,N. and Ooi,T. (1991) *Biochim. Biophys. Acta*, **871**, 45–54.
8. Margalit,H., Spouge,J.L., Cornette,J.L., Cease,K.B., Delisi,C. and Berzofsky,J.A. (1987) *J. Immunol.* **138**, 2213–2229.
9. Haynes,S.R., Mozer,B.A., Bhatia-Dey,N. and Dawid,I.B. (1989) *Dev. Biol.* **134**, 246–257.

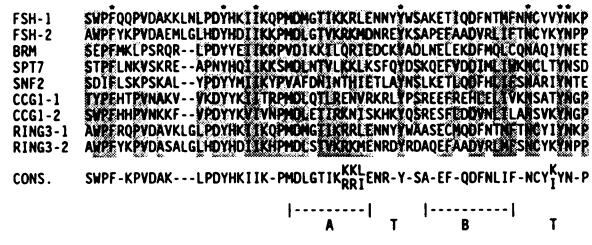

Figure 1. Sequences of the bromodomains and the derived consensus. Shaded residues indicate identities or conservative substitutions found in over half the examples. Asterisks mark the invariant residues. The consensus represents the most prevalent amino acid at each position (ie appearing at least three sequences) A and B mark the amphipathic helices; T indicates a reverse turn. Structure predictions were made using the programs PRDSEC (7) and AMPHI (8). Database accession numbers and references for sequences: FSH—M23221, M23222, ref. 9; BRM—M85049, ref. 1; SPT7—M87651; SNF2—M55906, ref. 4; CCG1—X07024, ref. 5; RING3—M80613, ref. 6.

*M23221, M23222, M85049, M87651, M55906, X07024 and M80613